



Genomic Loci Affecting Milk Production in German Black Pied Cattle (DSN)

Paula Korkuć¹, Danny Arends¹, Katharina May², Sven König² and Gudrun A. Brockmann^{1*}

¹ Albrecht Daniel Thaer-Institute for Agricultural and Horticultural Sciences, Animal Breeding Biology and Molecular Genetics, Humboldt University Berlin, Berlin, Germany, ² Institute of Animal Breeding and Genetics, Justus-Liebig-University of Giessen, Giessen, Germany

OPEN ACCESS

Edited by:

Ino Curik,
University of Zagreb, Croatia

Reviewed by:

Lingyang Xu,
Institute of Animal Sciences, Chinese
Academy of Agricultural Sciences,
China
Doreen Becker,
Leibniz Institute for Farm Animal
Biology (FBN), Germany

*Correspondence:

Gudrun A. Brockmann
gudrun.brockmann@hu-berlin.de

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 10 December 2020

Accepted: 11 February 2021

Published: 08 March 2021

Citation:

Korkuć P, Arends D, May K,
König S and Brockmann GA (2021)
Genomic Loci Affecting Milk
Production in German Black Pied
Cattle (DSN).
Front. Genet. 12:640039.
doi: 10.3389/fgene.2021.640039

German Black Pied cattle (DSN) is an endangered population of about 2,550 dual-purpose cattle in Germany. Having a milk yield of about 2,500 kg less than the predominant dairy breed Holstein, the preservation of DSN is supported by the German government and the EU. The identification of the genomic loci affecting milk production in DSN can provide a basis for selection decisions for genetic improvement of DSN in order to increase market chances through the improvement of milk yield. A genome-wide association analysis of 30 milk traits was conducted in different lactation periods and numbers. Association using multiple linear regression models in R was performed on 1,490 DSN cattle genotyped with BovineSNP50 SNP-chip. 41 significant and 20 suggestive SNPs affecting milk production traits in DSN were identified, as well as 15 additional SNPs for protein content which are less reliable due to high inflation. The most significant effects on milk yield in DSN were detected on chromosomes 1, 6, and 20. The region on chromosome 6 was located nearby the casein gene cluster and the corresponding haplotype overlapped the *CSN3* gene (casein kappa). Associations for fat and protein yield and content were also detected. High correlation between traits of the same lactation period or number led to some SNPs being significant for multiple investigated traits. Half of all identified SNPs have been reported in other studies, previously. 15 SNPs were associated with the same traits in other breeds. The other associated SNPs have been reported previously for traits such as exterior, health, meat and carcass, production, and reproduction traits. No association could be detected between *DGAT1* and other known milk genes with milk production traits despite the close relationship between DSN and Holstein. The results of this study confirmed that many SNPs identified in other breeds as associated with milk traits also affect milk traits in dual-purpose DSN cattle and can be used for further genetic analysis to identify genes and causal variants that affect milk production in DSN cattle.

Keywords: genome-wide association study, cattle, SNP chip, Holstein cattle, *CSN3* gene, *DGAT1* gene, casein gene

INTRODUCTION

German Black Pied cattle (DSN, “Deutsches Schwarzbuntes Niederungsrind”) is an endangered breed of about 2,550 dual-purpose cattle in Germany. The ancestor population of this breed is considered as one of the founder populations of the nowadays dominantly used high yielding German Holstein breed (Köppe-Forsthoff, 1967; Grothe, 1993), which is one reason why the German government and the European Union support its preservation. This support is necessary because milk yield of DSN cows is about 2,500 kg less compared to German Holstein, which has led to the replacement of DSN by Holstein cattle. Although DSN is a dual-purpose breed, milk yield is the main contributor to the economic merit and meat yield and carcass quality do not compensate the lower milk yield.

The interest in preservation of DSN does not only stem from its relationship to Holstein cattle and retaining genetic diversity as a gene reserve for the future. The interest in DSN cattle is also due to their advantageous traits. For example, the milk fat and protein content with 4.3 % and 3.7 %, respectively, is higher compared to German Holstein (RBB Rinderproduktion Berlin-Brandenburg GmbH, 2020). Moreover, DSN cattle are considered to be more robust for grazing and more fertile.

One of the long-term goals for maintaining DSN is to reduce governmental financial support by increasing economic value of the breed through the improvement of milk yield. Simultaneously, the advantageous traits and the typical body composition of this dual-purpose breed should be maintained.

So far, little is known about the genes affecting milk traits in DSN. Recently, genome-wide association studies (GWAS) in DSN for health traits identified three significant and two suggestive SNPs for clinical mastitis (Meier et al., 2020) and 44 significant SNPs for endoparasite resistance (May et al., 2019). Another study, investigating genomic variation in the casein gene cluster, found three protein variants of CSN2 and CSN3 and fixed protein variants for CSN1S1 and CSN1S2 (Meier et al., 2019). In contrast, 63,404 associations with milk traits in general were available from Cattle QTLdb Release 42 (accessed 09/21/2020) (Hu et al., 2019), whereof around 79 % were found in studies with Holstein cattle. Furthermore, 10 % of the reported associations in those studies with Holstein cattle are reported within the first 10 Mb on chromosome 14, where the *DGAT1* gene is located which is known to influence milk yield and composition (Grisart et al., 2002).

In this study, we investigated the genetic basis contributing to the variation in milk performance of the current DSN population by GWAS. Since the DSN population is small, power to find significant genomic loci is limited. Nevertheless, there is an urgent need to provide genetic association information for small endangered populations to support their preservation and further development. Even if not all genomic loci are significantly detectable, the obtained results in DSN and the comparison to related breeds provide a basis for selection decisions to genetically improve DSN.

MATERIALS AND METHODS

Populations

Data of 1,816 DSN cows was available for this study. These cows represent about two thirds of the DSN population registered in Germany (The Society for the Conservation of Old and Endangered Livestock Breeds (GEH), 2018). The cows were born between 2005 and 2016, and descended from 76 sires. Cows were raised on six farms. In order to reduce environmental influences, we filtered to have at least 20 DSN cows per farm, per sire, and per birth year. This reduced the data set to 1,490 DSN cows from five farms, born between 2007 and 2016, and descending from 28 sires.

In order to compare the results that we obtained in DSN with German Holstein, we used GWAS results previously obtained in our lab on a population of 2,400 German Holstein bulls. This population has been used and described in detail repeatedly (Zielke et al., 2011, 2013; Abdel-Shafy et al., 2018).

Phenotypes

Milk traits with corresponding pedigree data were obtained from the cattle breeding association “RBB Rinderproduktion Berlin-Brandenburg GmbH” in April 2020. Traits included milk, fat, and protein yield in kilogram (milk kg, fat kg, and protein kg) for three lactation periods: 100-days (100d), 200-days (200d), and 305-days (305d). 305 days data was available for the first three lactations (LA1, LA2, and LA3), whereas 100 and 200 days data was available only for LA1. 305 days data of cows with < 270 days in milk was not considered. Fat and protein content (fat %, protein %) were calculated by dividing fat or protein kg by milk kg of the respective lactation. The lactation mean (LAm) was calculated for cows with full data in the first three lactations. This leads to a total of 30 investigated milk traits in this study. For each trait, outliers were defined as values outside the 1.5 times interquartile range within each farm and removed from the data set. This leads to data being available for 1,478, 1,476, 1,372, 1,160, 862, and 685 DSN cows in LA1 (100d), LA1 (200d), LA1, LA2, LA3, and LAm, respectively.

Genotypes

DSN cattle cows were genotyped using the Illumina® Bovine50SNP v3 BeadChip (Illumina, Inc., 5200 Illumina Way, San Diego, CA, United States). SNP chip probe sequences were remapped against the *Bos taurus* genome version ARS-UCD1.2 (Rosen et al., 2018) using NCBI Nucleotide-Nucleotide BLAST version 2.2.31+ (Altschul et al., 1990) in order to obtain genome positions for SNPs on the ARS-UCD1.2 genome build. SNP probes that mapped to multiple genomic locations were removed. Genotype quality control was performed for animals and SNPs. SNP calls with a GC-score < 0.7 were set to missing. Animals with a call rate < 90% were discarded. SNPs with a call rate < 95% and a minor allele frequency (MAF) < 5% were removed. Lastly, genotype groups with less than 30 observations were set to missing to prevent spurious association. After quality control, 36,929 high confident SNPs were available for further analysis.

Genome-Wide Association Study

Genome-wide association studies was performed with multiple linear regression models implemented in the R language for statistical computing (version 4.0.3) (R Core Team, 2018). Various models were tested and compared by calculating the inflation factor λ of each model to judge the extent of the excess of type-I errors (Devlin and Roeder, 1999). Finally, the model with the lowest λ was selected (Supplementary Table 1). On average 27 % of top 100 SNPs of the selected model are shared with each of the other models tested (Supplementary Figure 1). The resulting model for testing the additive effect of each SNP was:

$$Y_{ijklmnopq} = ps_i + f_j^* + s_k^* + by_l^* + bs_m^* + cy_n^* + cs_o^* + ac_p^* + g_q + e_{ijklmnopq} \quad (1)$$

where $Y_{ijklmnopq}$ is the trait, ps_i represents the covariate for population stratification followed by the covariates for farm f_j , sire s_k , birth year by_l , birth season bs_m , calving year cy_n , calving season cs_o , age at first calving in days ac_p , the SNP genotype g_q , and $e_{ijklmnopq}$ is the residual error. Covariates marked with an asterisk “*” were only included into the model when the difference in the Akaike information criterion (ΔAIC) was ≤ -10 between the null model ($Y_i = ps_i$) and the null model extended with one of the covariates ($Y_{ix} = ps_i + covariate_x$) (Supplementary Table 2). All covariates were included as fixed effects as this resulted in the lowest inflation factor λ . Population stratification ps_i among the 1,490 DSN cows was examined using pairwise population concordance tests on an identity-by-state matrix implemented in PLINK (version 1.90) with a p -value cut-off of 0.0001 (Purcell et al., 2007; Chang et al., 2015). The resulting 33 clusters of relatedness were included as a covariate in the GWAS model.

Interactions between lactation stages and SNP genotypes were investigated for 100, 200, and 305 days performance data in LA1 and between lactation number and SNP genotypes for 305 days performance data in lactations 1 to 3. For testing the genetic interaction effects between milk, fat, and protein yield data and specific lactation periods, the difference between the difference between birth and 100 days, 100 and 200 days, and 200 and 305 days performance data was used as input data. A linear mixed-effects model as described by Lu and Bovenhuis (2019) was fitted using the R package lmerTest (version 3.1-3) (Kuznetsova et al., 2017):

$$Y_{ijklmnopqrs} = ps_i + f_j^* + s_k^* + by_l^* + bs_m^* + cy_n^* + cs_o^* + ac_p^* + g_q + L_r + g_q \times L_r + (1|animal_s) + e_{ijklmnopqrs} \quad (2)$$

where the same covariates were used as in Eq. 1 but extended by the lactation stage or lactation number L_r , the interaction term between lactation stage or lactation number and SNP genotype $L_r \times g_q$, and a random intercept included for each individual animal ($1|animal_s$) to compensate for repeated measurements on the animals. Covariates marked with an asterisk “*” were only included into the model when the difference in the Akaike information criterion (ΔAIC) was ≤ -10 between the

null model [$Y_s = (1|animal_s)$] and the null model extended with one of the covariates [$Y_{xs} = covariate_x + (1|animal_s)$] (Supplementary Table 3). Separate analyses were performed for the traits milk, fat and protein yield, and fat and protein content.

The significance threshold was adjusted for multiple testing using Bonferroni (BF) correction. The number of independent tests (M_{eff}) was estimated by the simpleM method, to account for linkage between neighboring SNPs (Gao et al., 2008). Window sizes from 100 to 630 (630 corresponds to the minimum number of SNPs in one out of all chromosomes) in this study were tested. The lowest estimated M_{eff} value was 18,278 at a window size of 630 which was then used for Bonferroni correction. After Bonferroni correction, SNPs were considered highly significant when $P_{BF} < 0.01$, significant when $P_{BF} < 0.05$, or suggestive when $P_{BF} < 0.1$. In the case of interaction terms, M_{eff} was multiplied by the number of lactation stages or lactation numbers ($n = 18,278 \times 3$) and p -values were noted as P_{BI} . Figures were produced using the R package ggplot2 (version 3.3.2) (Wickham et al., 2016) unless otherwise stated. For SNP effect plots, p -values between genotype groups were estimated using pairwise t -tests and displayed using R package ggpubr (version 0.4.0) (Kassambara, 2020).

Haplotypes and Gene Annotation

Haplotype blocks in which SNPs are located were computed using Haploview version 4.2 (Barrett et al., 2005). In order to define blocks in Haploview, the customized block definition was used with the option “solid LD spine” set to $D' > 0.6$. Genes within each haplotype were annotated using R package biomaRt (version 2.48.0) (Durinck et al., 2009) using the Ensembl *Bos taurus* database based on the ARS-UCD1.2 assembly (Yates et al., 2020). If no haplotype block could be estimated for a SNP, genes within ± 70 kb up- and down-stream (corresponding to the median haplotype block length of 140 kb) of the respective SNP position were considered as positional candidate genes. Additionally, the 1 Mb region centered around the investigated SNP was inspected for candidate genes. If consecutive 1 Mb regions overlapped, they were merged by taking the start of the first and the end of the second region.

Overlap With Cattle QTLdb

Associated SNPs were compared to GWAS results and known QTLs from Cattle QTLdb Release 42 (Hu et al., 2019) by using their SNP-IDs (rs). Further, also the 1 Mb region centered around the associated SNP was used to find overlapping loci with other publications. The entries from Cattle QTLdb were categorized into the trait categories “exterior,” “health,” “meat and carcass,” “milk,” “production,” and “reproduction.”

RESULTS

Associations for Almost All Traits

Genome-wide association analyses using 1,490 DSN cows and 36,929 high confident SNPs identified associations for all traits except for 100-days fat yield in lactation 1 (fat kg 100 days LA1) and the lactation mean of the 305-days fat yield (fat kg LAM). The

inflation factor λ ranged from 1.27 for fat yield (LA2 305 days) to 2.19 for protein content (LA2 305 days) although extensive efforts were taken to select a GWAS model that showed the lowest overall inflation (**Supplementary Table 1**). Nonetheless, the inflation factor λ and the Q-Q-plots for all traits (**Supplementary Figure 2**) showed that the level of statistical significance was overestimated in general. We assume that this especially holds for protein content in all lactation periods and numbers and for fat content in LA1 (100 and 305 days). The number of false positives was increased as λ -values higher or equal to 1.5 were observed for these traits. Even if we reduce the genome-wide significance threshold to $P_{BF} < 0.0005$, 16 SNPs remain associated. The results of those traits are listed in the **Supplementary Table 4**. Results of associations for 20 traits where λ was below 1.5 are presented in **Table 1**. For those traits, 14 highly significant, 27 significant, and 20 suggestive SNPs were found. Some SNPs were associated with multiple traits because of high correlation between those traits (**Supplementary Tables 5,6**). High correlation was found between performance data within the first lactation (100, 200, and 305 days of LA1) ($r > 0.76$), and between traits of the first three lactations (LA1, LA2, and LA3) and the lactation mean (LAM) ($r > 0.77$). Fat and protein yield within a lactation were also highly correlated ($r > 0.75$).

Effects on Milk Yield

Since milk yield is the economically most important production trait, there is major interest in identifying genetic loci contributing to its variance in DSN cows (**Table 1** and **Figure 1**). Using the mean milk yield across the first three lactations (LAM), only two loci were found to be associated, one on chromosome 8 at 53.7 Mb (8:53,663,120, rs41793393, $P_{BF} = 0.03554$) and another one on chromosome 9 at 10.6 Mb (9: 10,638,013, rs133869947, $P_{BF} = 0.04397$) (**Supplementary Figures 3A,B**). These two loci were also associated with milk yield in lactation 3 (LA3). Examining different lactation periods and lactations separately, we found additional loci on chromosomes 1, 4, 5, 6, 20, 24, 25, and X. Since the significance varied largely between lactation periods and lactation numbers, the interaction between the SNPs and lactation stage or lactation number was also investigated (**Supplementary Table 7**).

In the first lactation (LA1), chromosome 6 was most significantly associated (**Table 1** and **Figure 1**). In the region between 60.2 and 87.2 Mb, 9 SNPs were associated with milk yield in all lactation periods of LA1. The top marker rs109592101 (6:86,112,142, $P_{BF} = 0.00009$, **Supplementary Figure 3C**) showed the highest association with the 100 days performance. The minor allele A with a frequency of 0.42 accounted for a decrease in milk yield of 78 kg, 119 kg and 193 kg after 100, 200, and 305 days in lactation, respectively. The same SNP also suggestively affected ($P_{BF} = 0.08431$) protein yield with a minor allele effect of -1.8 kg after 100 days in lactation. In the haplotype block around the top SNP (6:85,633,295-87,011,619, **Figure 2** and **Supplementary Table 8**) 16 genes are located, among them CSN3 (kappa casein) is known as main milk protein gene and as a gene affecting milk yield and composition in Holstein cattle (Mckenzie et al., 1984; Ng-Kwai-Hang et al., 1984). Another SNP in the same region on chromosome 6 (rs110291935, 6:80,530,130,

$P_{BF} < 0.02$, **Supplementary Figure 3D**) decreased milk yield until 100, 200, and 305 days in LA1 by 71, 132, and 209 kg, respectively. The corresponding haplotype block (6:80,530,130-80,626,467) did not contain any gene (**Supplementary Table 8**). However, the surrounding 1 Mb region harbors three genes ENSBTAG00000050977, *EPHA5* (EPH receptor A5), and ENSBTAG00000053125. Additional SNPs for milk yield in LA1 were found on chromosome 1 (rs42347234, 1:115,539,332, $P_{BF} = 0.03846$; rs41255272, 1:143,812,919, $P_{BF} = 0.01478$), chromosome 4 (rs42753220, 4:91,071,208, $P_{BF} = 0.01390$), chromosome 5 (rs41652414, 5:114,935,739, $P_{BF} = 0.02814$), and chromosome 25 (rs109027867, 25:11,019,450, $P_{BF} = 0.03396$) (**Supplementary Figures 3E-I**). The minor allele of these SNPs showed an increase in milk yield by at least 64 kg, 141, and 204 kg for 100, 200, and 305 days performance data of LA1, respectively, except for the SNP rs109027867 on chromosome 25 which showed a decrease of the minor allele by 128 kg for 200 days data.

In lactation 2 (LA2), the only association with milk yield was found on chromosome 20 (**Table 1** and **Figure 1**) with the SNP rs110353352 (20:71,448,297, $P_{BF} = 0.00040$, **Supplementary Figure 3J**). The minor allele C, which segregated at a frequency of 0.13, was advantageous for milk production increasing milk yield by 484 kg. This same SNP allele had also positive effects of 17.3 kg and 15.0 kg on fat and protein yield in LA2, respectively. The haplotype block of this SNP (20:71,378,297-71,518,297) contains five genes *TRIP13* (thyroid hormone receptor interactor 13), *BRD9* (bromodomain containing 9), ENSBTAG00000054687, *TPPP* (tubulin polymerization promoting protein), and *CEP72* (centrosomal protein 72) (**Supplementary Table 8**).

In lactation 3 (LA3), the most significant association for milk yield was found on chromosome 1 (**Table 1** and **Figure 1**). The top SNP rs43246393 (1:79,757,250, $P_{BF} = 0.00020$, **Supplementary Figure 3K**) was also associated with fat ($P_{BF} = 0.01868$) and protein yield ($P_{BF} = 0.00007$) in LA3 as well as with protein yield of the average across all three lactations ($P_{BF} = 0.02641$). The minor allele T, segregating at a frequency of 0.34, accounted for 390 kg, 13.4 kg and 13.2 kg less milk, fat, and protein in LA3, respectively, and for 10.3 kg less protein in the lactation mean (LAM). The test for interaction between genotypes at this top SNP and lactation stage or lactation number showed significant effects on milk, fat, and protein yields ($P_{BI} < 0.03$, **Supplementary Table 7**). The corresponding haplotype block around this SNP (1:79,606,618-79,757,250) contains three genes *BCL6* (BCL6 transcription repressor), *RTP2* (receptor transporter protein 2), and *SST* (somatostatin). In addition to the locus on chromosome 1, two SNPs on chromosome 8 were found to be associated (rs41793393, 8:53,663,120, $P_{BF} = 0.00508$, **Supplementary Figure 3L**; rs109542652, 8:53,867,972, $P_{BF} = 0.02538$). As mentioned above, this region was also identified for the average milk yield over lactations 1–3 (LAM). The minor allele T of the lead SNP rs41793393 (frequency 0.19) is decreasing milk yield in LA3 by 356 kg and in LAM by 316 kg. The corresponding haplotype block (8:53,467,317-54,449,954) contains the genes *GNA14* (G protein subunit alpha 14), *GNAQ* (G protein subunit alpha q), *CEP78* (centrosomal protein 78), and *PSAT1* (phosphoserine aminotransferase 1). Additional

TABLE 1 | GWAS for milk production traits in DSN.

Chr	Position (bp)	SNP-ID	Ref	Alt	MA	MAF	Trait	N	β	SE(β)	P _{BF}
1	77,049,090	rs110516247	A	C	C	0.11	Milk kg (LA3)	846	−515	100	0.00679
							Protein kg (LA3)	844	−16.3	3.3	0.02351
	79,757,250	rs43246393	C	T	T	0.34	Milk kg (LA3)	856	−390	76	0.00020
							Protein kg (LA3)	854	−13.2	2.5	0.00007
							Protein kg (LAm)	679	−10.3	2.2	0.02641
							Fat kg (LA3)	857	−13.4	3.1	0.01868
	115,539,332	rs42347234	C	T	T	0.11	Milk kg (LA1)	1355	336	70	0.03846
							Protein kg (LA1)	1357	11.9	2.4	0.01140
	128,503,005	rs109686415	C	T	C	0.49	Fat % (LA2)	1115	−0.082	0.016	0.00357
							Fat % (LAm)	660	−0.082	0.016	0.01599
	143,812,919	rs41255272	A	G	A	0.50	Milk kg (LA1)	1367	204	41	0.01478
							Protein kg (LA1)	1369	6.6	1.4	0.04613
2	123,100,345	rs110278850	A	C	C	0.45	Protein kg (200 days)	1436	3.5	0.8	0.04941
3	15,947,663	rs110565504	T	C	T	0.24	Fat % (200 days)	1458	0.056	0.019	0.03398
4	47,674,986	rs110143001	G	A	G	0.42	Fat % (LA3)	844	0.086	0.019	0.04275
	91,071,208	rs42753220	C	T	C	0.40	Milk kg (LA1)	1370	213	44	0.01390
5	62,094,191	rs29009717	A	G	A	0.18	Fat % (LA3)	853	−0.125	0.032	0.09655
	83,959,138	rs41660560	G	T	G	0.35	Protein kg (LA2)	1160	9.4	2.1	0.01767
	93,953,629	rs109945272	T	C	T	0.24	Fat % (LA2)	1136	−0.077	0.024	0.08780
	114,935,739	rs41652414	C	T	T	0.19	Milk kg (LA1)	1357	220	71	0.02814
							Milk kg (100 days)	1457	64	24	0.08596
							Milk kg (200 days)	1452	141	46	0.00923
							Protein kg (LA1)	1359	7.2	2.4	0.04450
							Protein kg (100 days)	1459	2	0.7	0.08697
							Protein kg (200 days)	1457	4.1	1.4	0.02288
							Fat kg (200 days)	1447	5.1	1.7	0.07535
6	60,162,206	rs41605188	G	A	A	0.34	Milk kg (200 days)	1469	119	29	0.09377
	61,733,082	rs135571989	C	A	A	0.21	Milk kg (LA1)	1320	−195	71	0.07717
							Milk kg (200 days)	1416	−110	45	0.08954
	62,988,117	rs42436495	T	G	G	0.30	Milk kg (LA1)	1370	225	49	0.01789
							Milk kg (200 days)	1469	136	32	0.03921
	63,010,380	rs42436482	A	G	G	0.29	Milk kg (LA1)	1368	220	52	0.04884
							Milk kg (200 days)	1467	132	34	0.09298
	64,928,624	rs42482917	T	C	C	0.32	Milk kg (200 days)	1463	111	30	0.09593
	77,688,509	rs41652041	A	G	G	0.35	Milk kg (LA1)	1372	−212	49	0.06866
							Milk kg (200 days)	1470	−138	31	0.03904
	80,530,130	rs110291935	T	C	T	0.41	Milk kg (LA1)	1370	−209	42	0.01362
							Milk kg (100 days)	1474	−71	14	0.00519
							Milk kg (200 days)	1469	−132	27	0.00655
	80,626,467	rs109872424	T	C	T	0.33	Milk kg (100 days)	1473	−64	16	0.06727
							Milk kg (200 days)	1468	−124	30	0.02416
	86,112,142	rs109592101	A	G	A	0.42	Milk kg (LA1)	1372	−193	41	0.01990
							Milk kg (100 days)	1476	−78	14	0.00009
							Milk kg (200 days)	1471	−119	26	0.04162
							Protein kg (100 days)	1478	−1.8	0.4	0.08431
	87,266,808	rs41591365	C	T	C	0.46	Protein kg (100 days)	1473	2.1	0.4	0.03608
	88,164,411	rs41622837	A	G	A	0.14	Protein kg (200 days)	1444	−6	1.3	0.07033
8	53,663,120	rs41793393	C	T	T	0.19	Milk kg (LA3)	847	−356	115	0.00508
							Milk kg (LAm)	673	−316	97	0.03554
	53,867,972	rs109542652	G	A	A	0.14	Milk kg (LA3)	846	−489	101	0.02538
	59,101,606	rs43550935	A	G	G	0.19	Milk kg (LAm)	685	−271	95	0.07475
							Protein kg (LAm)	685	−9.1	3.2	0.02661
	100,876,785	rs108983661	G	A	A	0.25	Fat kg (LAm)	683	12.5	3	0.05697

(Continued)

TABLE 1 | Continued

Chr	Position (bp)	SNP-ID	Ref	Alt	MA	MAF	Trait	N	β	SE(β)	P_{BF}
9	10,638,013	rs133869947	A	C	A	0.38	Milk kg (LA3)	818	−297	70	0.06280
							Milk kg (LAm)	645	−270	60	0.04397
10	26,353,699	rs110314239	G	T	T	0.48	Fat kg (LA1)	1366	−6.8	1.5	0.05814
							Fat % (200 days)	1449	−0.061	0.013	0.03506
11	69,443,503	rs110564084	T	C	T	0.30	Protein kg (LA2)	1157	11.9	2.5	0.05681
							Fat % (200 days)	1454	0.057	0.013	0.09715
12	66,340,756	rs41629344	T	G	G	0.39	Fat % (200 days)	1452	0.06	0.013	0.00380
							Protein kg (100 days)	1470	4	0.9	0.06447
14	26,340,400	rs41727315	G	A	A	0.06	Fat % (200 days)	1459	−0.082	0.02	0.05676
							Fat % (LA3)	835	0.153	0.031	0.02338
16	30,668,830	rs43041491	G	T	G	0.11	Fat % (200 days)	1458	−0.055	0.015	0.05996
							Fat % (LA2)	1134	−0.061	0.015	0.02776
	33,789,714	rs41796289	C	T	C	0.39	Fat % (200 days)	1450	−0.049	0.013	0.09433
							Fat % (200 days)	1446	0.058	0.012	0.04642
	40,391,486	rs41804404	A	G	G	0.46	Protein kg (100 days)	1444	−1.8	0.5	0.05105
							Protein kg (100 days)	1472	1.9	0.4	0.04084
18	53,596,284	rs109907036	C	T	T	0.37	Fat % (LA2)	1136	0.086	0.017	0.00477
							Fat % (200 days)	1445	−0.118	0.023	0.00777
20	50,879,180	rs41948928	T	C	T	0.10	Milk kg (LA2)	1137	484	86	0.00040
							Protein kg (LA2)	1143	15	2.9	0.00503
	71,448,297	rs110353352	C	T	C	0.13	Fat kg (LA2)	1137	17.3	3.5	0.01434
							Protein kg (200 days)	1450	−4.2	0.9	0.09136
21	42,828,439	rs41978846	T	G	T	0.33	Milk kg (LA3)	862	−348	72	0.01192
							Protein kg (LA1)	1355	−7	1.5	0.01207
24	16,789,760	rs110860585	C	T	T	0.41	Protein kg (200 days)	1453	−4.1	0.9	0.03475
							Fat kg (LA1)	1348	−8.5	1.7	0.02349
25	7,944,597	rs109583598	T	C	C	0.29	Fat % (200 days)	1456	0.065	0.016	0.01082
							Fat kg (LA1)	1368	10.8	3.5	0.00392
	9,711,895	rs110469759	A	G	A	0.17	Fat kg (200 days)	1467	5.7	2	0.02959
							Milk kg (200 days)	1445	−128	26	0.03396
27	13,864,569	rs110009442	C	G	C	0.32	Fat % (LA2)	1125	0.061	0.018	0.08067
							Fat % (200 days)	1442	0.06	0.015	0.09215
	18,527,112	rs42957103	G	A	A	0.41	Fat kg (LA1)	1367	7.5	1.7	0.09536
							Protein kg (100 days)	1419	3.8	0.8	0.06929
28	25,196,334	rs41587054	A	G	G	0.10	Fat % (LA2)	1137	0.076	0.016	0.04229
							Fat % (LA2)	1118	−0.082	0.016	0.00338
29	50,217,955	rs109241029	G	A	A	0.44	Fat % (LA2)	1113	−0.09	0.016	0.00022
							Fat % (LA2)	1113	−0.09	0.016	0.00022
X	12,896,716	rs109188619	T	A	T	0.18	Milk kg (LA3)	860	546	122	0.02488
							Protein kg (200 days)	1476	−4.6	1.1	0.08409
	116,886,837	rs41626783	T	C	C	0.38	Protein kg (LA3)	851	−13.5	2.7	0.01321
							Protein kg (LAm)	676	−10.8	2.4	0.07065
	117,691,901	rs110011913	A	T	T	0.31	Fat kg (200 days)	1409	−6.8	1.4	0.07242

For each trait, significantly associated SNPs are listed with chromosome number (Chr), chromosomal position in base pairs (bp), the SNP reference ID (SNP-ID), the given allele on the forward strand in the reference genome ARS-UCD1.2 (Ref), the alternative allele (Alt), the minor allele in the examined DSN population (MA), and the minor allele frequency (MAF). Furthermore, the associated trait, the number of cows in the specific analysis of this trait (n), the allele substitution effect of the minor allele (β), the standard error [SE(β)], and the p-value after Bonferroni-correction (P_{BF}) are given. P-values of $P_{BF} < 0.1$, $P_{BF} < 0.05$, and $P_{BF} < 0.01$ were considered as suggestive (highlighted in gray), significant, or highly significant (highlighted in bold), respectively.

associations with milk yield in LA3 were found for SNP rs110860585 (24:16,789,760, $P_{BF} = 0.01192$, **Supplementary Figure 3M**) and rs109188619 (X:12,896,716, $P_{BF} = 0.02488$, **Supplementary Figure 3N**).

Effects on Milk Fat Yield and Content

Regions associated with milk fat yield were identified on chromosomes 1, 20, 24, and 25 (Table 1 and **Supplementary**

Figure 4). The most significant effect on fat yield was found on chromosome 25. The SNP rs110469759 (25:9,711,895, $P_{BF} = 0.00392$, **Supplementary Figure 3O**) with an allele frequency of 0.17 showed an allele substitution effect of the A allele leading to an increase of 10.8 kg fat in LA1 and of 5.7 kg for 200 days performance in LA1. The association for fat yield on the other chromosomes (LA3: rs43246393, 1:79,757,250, $P_{BF} = 0.01868$; LA2: rs110353352, 20:71,448,297, $P_{BF} = 0.01434$;

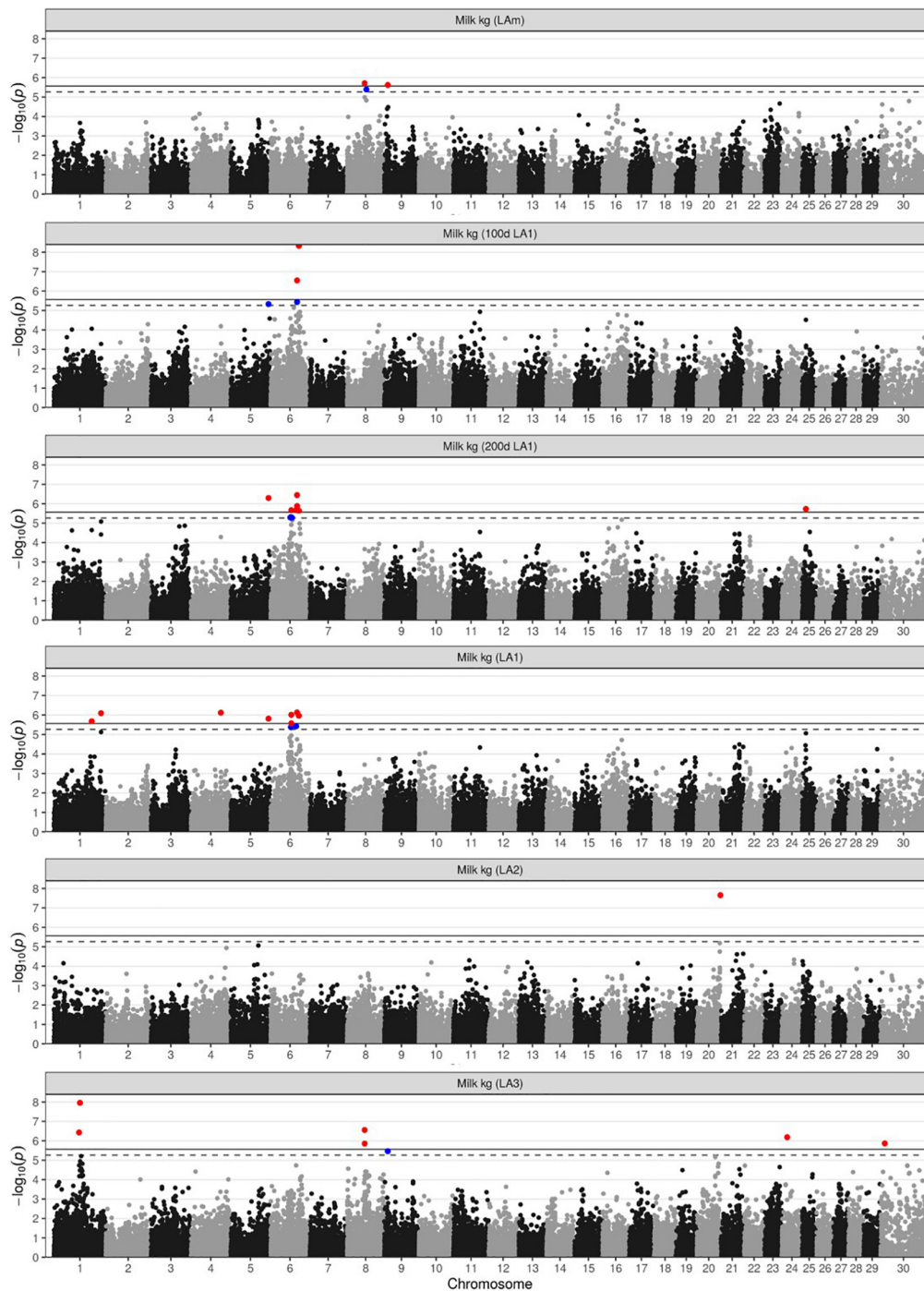


FIGURE 1 | Manhattan plots for milk yield in kg. Plots are shown for the lactation mean of the 305 days performance (LAM), the 100 and 200 days performance in LA1, and the 305-days performance in the first three lactations (LA1-LA3). Markers above the significance or suggestive thresholds are highlighted in red (solid line, $\alpha < 0.05$) or blue (dashed line, $\alpha < 0.1$), respectively.

LA1: rs110476141, 24:30,250,034, $P_{BF} = 0.02349$, **Supplementary Figures 3P-R**) coincided with effects on milk and protein yield. The minor alleles on chromosomes 1 and 24 accounted for lower, the minor allele on chromosome 20 and 25 for higher milk, protein and fat yields (**Table 1**).

Highly significant SNPs associated with fat content were identified on chromosomes 1, 12, 18, 20, and 29 (**Table 1** and **Supplementary Figure 5**). The most significant association was found on chromosome 29 with the top SNP rs110740589 (29:50,326,170, $P_{BF} = 0.00022$, **Supplementary**

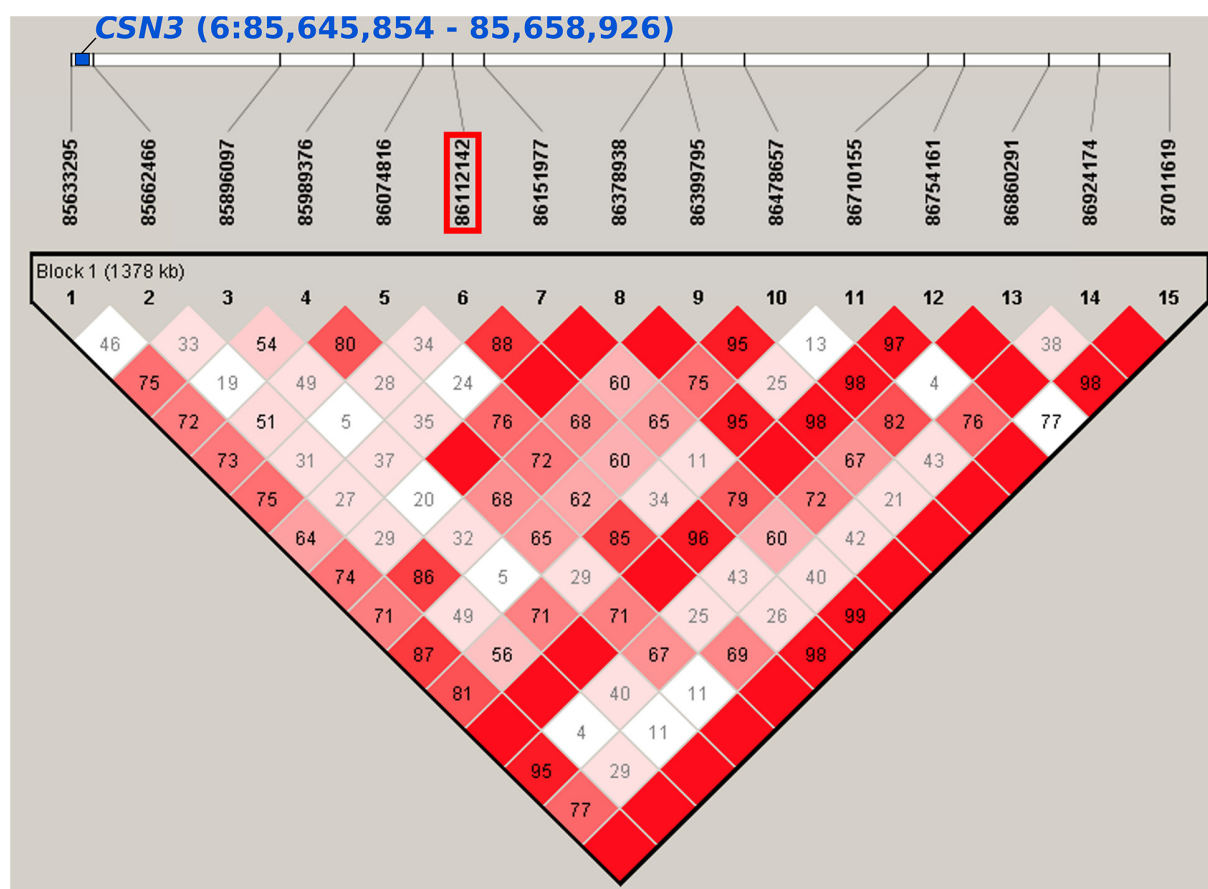


FIGURE 2 | Haplotype block including SNP rs109592101 (6:86,112,142) estimated with Haploview. The investigated SNP is located in a haplotype block (6:85,633,295-87,011,619) that overlaps with the position of the *CSN3* gene (6:85,645,854-85,658,926) which is known to be associated with milk production traits in several cattle breeds.

Figure 3S). This SNP was associated with fat content in LA2. The minor allele A of this SNP (MAF = 0.41) showed a decrease in fat content in LA2 by 0.09% points. The neighboring SNP rs109840529 (29:50,260,533) showed interaction effects with the 100–305 days lactation stages of LA1 on fat content ($P_{BI} = 1.4E-07$, **Supplementary Table 7**). On chromosome 1, the SNP rs109686415 (1:128,503,005) was associated with fat content in LA2 ($P_{BF} = 0.00357$, **Supplementary Figure 3T**) and with the lactation mean of LA1-3 ($P_{BF} = 0.01599$). This SNP was also significant for the interaction effect between SNP and lactation stages for the trait fat content across ($P_{BI} = 0.00016$, **Supplementary Table 7**). The SNP rs109907036 on chromosome 18 (18:53,596,284, $P_{BF} = 0.00477$, **Supplementary Figure 3W**) was also identified for fat content in LA2. The associations on chromosomes 12 (rs41629344, 12:66,340,756, $P_{BF} = 0.00380$, **Supplementary Figure 3U**) and 20 (rs41948928, 20:50,879,180, $P_{BF} = 0.00777$, **Supplementary Figure 3V**) were associated with fat content for 200 days performance data in LA1. The minor allele was the non-beneficial allele on chromosome 1 and 20, and the beneficial allele on chromosomes 12 and 18.

Additional loci affecting the changing fat contents during lactation 1 were found on chromosomes 5 and 6 ($p < 0.0006$, **Supplementary Table 7**). On chromosome 5 these were the SNPs rs41660560 (5:83,959,138, $P_{BI} = 0.00003$) and rs109945272 (5:93,953,629, $P_{BI} = 1.9E-06$) and on chromosome 6 the four SNPs rs41605188 (6:60,162,206, $P_{BI} = 1.1E-07$), rs42436495 (6:62,988,117, $P_{BI} = 0.00005$), rs42436482 (6:63,010,380, $P_{BI} = 1.7E-07$), and rs42482917 (6:64,928,624, $P_{BI} = 2.3E-12$). The changing effects of fat content of these loci were significant only when examining the interaction between SNP and lactation stage in LA1.

For the most significant region on chromosome 29, the haplotype block (29:50,229,562-50,326,170) harboring the two SNPs rs109840529 and rs110740589 contained one gene of unknown function (ENSBTAG00000050398) (**Supplementary Table 8**). The SNP on chromosome 1 was located in a haplotype block (1:128,226,756-128,622,561) harboring the genes *TRIM42* (tripartite motif containing 42) and *CLSTN2* (calsynenin 2). The haplotype block around the SNP on chromosome 18 (18:53,340,459-53,596,284) included 14 genes, among them *FOXA4* (forkhead box A3) and *IGFL1* (IGF like family member 1).

Effects on Milk Protein Yield and Content

Highly significant effects ($P_{BF} < 0.01$) on milk protein yield were identified on chromosomes 1 and 20 with data from LA3 and LA2, respectively (Table 1 and Supplementary Figure 6). The SNPs rs43246393 (1:79,757,250, $P_{BF} = 0.00007$, Supplementary Figure 3X) and rs110353352 (20:71,448,297, $P_{BF} = 0.00503$, Supplementary Figure 3Y) on these two chromosomes showed besides their effects on milk protein yield also an effect on milk and fat yield in the same lactations. While the minor allele of the chromosome 1 SNP was disadvantageous for all yield traits, the minor allele of the chromosome 20 SNP was favorable.

All association tests with milk protein content resulted in highly inflated p -values with $\lambda \geq 1.5$ (Supplementary Figure 2). Therefore, we decided to focus on most significant results by lowering the genome-wide significance threshold 100-fold from $P_{BF} < 0.05$ to $P_{BF} < 0.0005$. This resulted in 16 associated SNPs for protein content in different lactation periods and numbers (Supplementary Table 4 and Supplementary Figure 7). These associations were found on chromosomes 2, 3, 5, 6, 10, 18, 20, and 28. The most significant association was found on chromosome 3 (rs110474631, 3:21,764,453, $P_{BF} = 4.5E-06$, Supplementary Figure 3Z). This SNP was found for protein content in LA1 (200, 305 days), LA2, and LAm. The minor allele G of this SNP had a frequency of 0.20 and decreased protein content by on average 0.05 % points in the before mentioned lactation periods and numbers. The corresponding haplotype block (3:21,764,453–21,819,709) does not contain any gene, but the 1 Mb region harbors 23 different genes (Supplementary Table 9). The SNP rs110291935 located on chromosome 6 was the second highest association found in all periods of LA1 (6:80,530,130, $P_{BF} < 0.0003$, Supplementary Figure 3A2). The minor allele T of this SNP increased protein content by 0.045–0.047 % points. The same SNP was already found to be associated with milk yield in all periods of LA1 (Table 1), where the minor allele decreased milk yield by 71, 132, and 209 kg in the 100, 200, and 305 days performance data, respectively. The corresponding haplotype and candidate genes were described in detail in the section of associations with milk yield. The third highest association was found on chromosome 10 at 46.5 Mb with the top SNP rs109605174 (10:46,450,562, $P_{BF} = 6.5E-06$, Supplementary Figure 3B2) in LA2. The minor allele of this SNP increased protein content by 0.055 % points. The neighboring SNPs rs109277788 (10:44,773,979, $P_{BF} = 0.00002$) and rs43625129 (10:47,670,717, $P_{BF} = 9.2E-06$) increased also protein content in LA2 and in the first 100 days in LA1. The corresponding haplotype block (10:46,330,098–46,672,954) comprises seven genes ENSBTAG00000054388, ENSBTAG00000050908, ENSBTAG00000019474, *FBXL22* (F-box and leucine rich repeat protein 22), *USP3* (ubiquitin specific peptidase 3), *CA12* (carbonic anhydrase 12), and *APH1B* (aph-1 homolog B, gamma-secretase subunit).

Missing Associations to Regions With Known Effects on Milk Production

Since some genes are well known for their effects on milk yield and composition in Holstein cattle, we separately tested

the association of SNPs in close proximity (<500 kb) to such candidate genes (Ogorevc et al., 2009). We found 19 SNPs on the SNP chip that were in close proximity to 10 candidate genes for milk production (*ABCG2*, *CSN1S1*, *CSN1S2*, *CSN2*, *DGAT1*, *GHR*, *GPAT4*, *PAEP*, *PRLR*, and *SPP1*). None of those 19 SNPs were significantly or suggestively associated with any of the investigated traits in DSN (Supplementary Table 10 and Supplementary Figure 8).

Overlap With Other Publications

Overall 31 out of 76 identified SNPs (including 16 highly significant SNPs associated with protein content) that were found to be associated with milk traits in DSN were reported in other studies (Supplementary Table 11) whose results were available in the Cattle QTLdb. Seven SNPs were found to be associated with the same milk traits in Holstein cattle as in DSN. These SNPs were located on chromosome 3 at 15 Mb (rs110073735, 3:15,470,670) and 22 Mb (rs41587408, 3:21,692,628), chromosome 6 between 86–87 Mb (rs109592101, 6:86,112,142; rs41591365, 6:87,266,808), chromosome 10 at 46 Mb (rs109605174, 10:46,450,562), chromosome 20 at 51 Mb (rs41948928, 20:50,879,180), and chromosome X at 13 Mb (rs109188619, X:12,896,716) (Kolbehdari et al., 2009; Cole et al., 2011; Meredith et al., 2012; Nayeri et al., 2016; Wang and Chatterjee, 2017; Jiang et al., 2019). Eight additional SNPs that were significant in DSN were found to be associated with other milk traits in Holstein and Brown Swiss cattle. Those were located on chromosome 5 at 75 Mb (rs110803736, 5:74,853,402), chromosome 6 between 65–81 Mb (rs42482917, 6:64,928,624; rs42224984, 6:65,962,733; rs41652041, 6:77,688,509; rs110291935, 6:80,530,130; rs109872424, 6:80,626,467), chromosome 8 at 54 Mb (rs41793393, 8:53,663,120), and chromosome 16 at 40 Mb (rs41804404, 16:40,391,486) (Poulsen et al., 2015; Buitenhuis et al., 2016; Dadousis et al., 2017; Jiang et al., 2019). Further, 16 associations to exterior, health, meat and carcass, production, reproduction traits in diverse breeds were reported (Snelling et al., 2010; Bolormaa et al., 2011; Cole et al., 2011; Hawken et al., 2012; McClure et al., 2012; Doran et al., 2014; Aliloo et al., 2015; Mészáros et al., 2015; Mapholi et al., 2016; Parker Gaddis et al., 2016; Li et al., 2017; Mateescu et al., 2017; Nayeri et al., 2019). When expanding the region of interest not only to single SNPs reported in this study, but to 1 Mb regions centered at the respective SNPs, all of the regions identified in DSN overlap with associations uploaded to the Cattle QTLdb.

Especially, for the region on chromosome 6 between 80.5–87.2 Mb (flanking SNPs: rs110291935 and rs41591365), which was significantly associated with milk and protein yield in all investigated periods of LA1 in our study, many different associations could be found in Cattle QTLdb. These associations included milk and protein yield as well as milk kappa-casein content in Holstein cattle (Meredith et al., 2012; Buitenhuis et al., 2016; Jiang et al., 2019), curd firmness and cheese fat recovery in Brown Swiss cattle (Dadousis et al., 2017), health traits such as somatic cell score in Holstein cattle (Jiang et al., 2019), and exterior traits such as facial pigmentation in Fleckvieh cattle (Mészáros et al., 2015).

DISCUSSION

In this study, we investigated genetic factors underlying milk production traits in DSN cattle. We detected 41 significant and 20 suggestive SNPs mostly found in regions previously associated to milk production traits or other trait categories. The high overlap of associated SNPs in this study with other studies shows that the regions that influence milk traits in DSN may have similar or even pleiotropic functions in other breeds.

We observed a high overlap of identified genomic regions in DSN and Holstein, even if we did not find significant effects of major candidate genes that affect milk production in Holstein cattle. While the key genes driving e.g., milk yield or fat and protein content might be different between DSN and Holstein, it is surprising that especially the well described association in the region of and around *DGAT1* was not detected in DSN although DSN and Holstein cattle are closely related. Although the MAFs of the three closest SNPs to *DGAT1* gene are high (MAF = 0.44–0.47) in DSN (**Supplementary Table 10**), we observed that the actual variant causing the A232K substitution is very rare having a MAF of 0.02 (preliminary results of sequencing data from 57 DSN cattle). It is possible that not only the causal variant of *DGAT1*, but also of other known milk genes are rare in DSN and thus do not have much effect in the DSN population.

Several interesting candidate genes were found including *CSN3* gene (kappa casein) that belongs to the casein cluster on chromosome 6 at 86 Mb, which are known to influence milk fat and protein content as well as milk properties (Caroli et al., 2009), also plays an important role in DSN for milk and protein yield. The genetic architecture of the casein gene cluster of DSN in comparison to other breeds was recently investigated in detail (Meier et al., 2019). Among 14 investigated breeds in that study, the casein gene cluster of DSN was most similar to Danish Red. One of the most significant associations on chromosome 1 at 79 Mb was close to *SST* (somatostatin) and *BCL6* (*BCL6* transcription repressor). Since *SST* regulates the secretion of pituitary hormones including prolactin (Eigler and Ben-Shlomo, 2014), it may contribute to the regulation of milk production (Dybus, 2002; Alipanah et al., 2007). And *BCL6* was found to be expressed in mammary epithelium and also in breast cancer (Logarajah et al., 2003). Also interesting is *SLC6A3* gene (solute carrier family 6 member 3) on chromosome 20 at 71 Mb to which the Gene Ontology term “lactation” (GO:0007595) was assigned. This assignment was inferred from electronic annotation by Gene Ontology Consortium (Ashburner et al., 2000) and is possibly based on the homology of this gene in mice where it was shown that mutations in this gene cause lactation failure (Bossé et al., 1997). Beside *SLC6A3*, other solute carrier family genes were found in the same region: *SLC6A18* (family 6 member 18), *SLC6A19* (family 6 member 19), *SLC9A3* (family 9 member 3), and *SLC12A7* (family 12 member 7). Also, genes coding for enzymes or receptors targeting milk ingredients such as prostaglandin and riboflavin were found, e.g., *PTGR1* (prostaglandin reductase 1) on chromosome 8 at 101 Mb, *PTGIR* (prostaglandin I2 receptor) on chromosome 18 at 54 Mb, and *FLAD1* (flavin adenine dinucleotide synthetase 1) on chromosome 3 at 15 Mb.

Since population stratification has a significant effect on the number of false positive QTLs detected during association analysis, we tried multiple approaches to reduce the inflation factor λ for different traits of interest. The higher the inflation factor λ , the higher the number of false-positively reported significant associations in GWAS analyses. Although we tried to compensate for population stratification by using pairwise population concordance test and paternal pedigree information, we were unable to capture the population structure entirely. That is most likely due to the fact that the population size of DSN is small, and the number of breeding bulls used each generation is limited. As such, DSN animals are not unrelated and complex family relationships exist between animals, which is a common issue in small populations. Thus, we assume that the residual inflation is genetically caused due to high linkage disequilibrium between SNPs. Moreover, using both population stratification and pedigree information led to an overfitting of the model that took away genetic variance. Further genetic variance was lost as we required that each sire was represented by at least 20 offspring in the phenotype data and thus the number of sires contributing to the analyzed DSN population was reduced by two-thirds. By doing so, also the number of animals available for GWAS was reduced and as a result the statistical power to detect significant associations. Nonetheless, we believe that lowering the genetic variance and statistical power in favor of lower inflation of p -values is the proper way to prevent false positive associations.

The results of this study could be used as prior information to detect SNPs associated with milk production traits in other (related) breeds. For that, cattle from other breeds could be genotyped for single SNPs (e.g., top SNPs) instead of being genotyped using SNP chips or even being whole genome sequenced which would reduce costs. Furthermore, since genotypic data of a sufficient amount of DSN cattle was not available yet, genomic breeding is not yet established. The identified SNPs and their associations could be directly used to make breeding decisions by DSN breeders. Even predictions about performance of young DSN cows that do not have any phenotypic data yet could be made if genotypic data would be available. Thus, a first step toward genomic breeding for milk traits in DSN cattle was made. Further analyses are needed to evaluate whether these associated SNPs do not show any negative effect on other trait categories such as fertility.

CONCLUSION

The genome-wide association analysis identified several significant regions associated with milk traits in diverse lactation periods and numbers in dual-purpose DSN cattle. In the view of the fact that the biggest possible sample size of the investigated DSN population was still relatively small, the most promising SNPs for improving milk production in DSN are those that were also significantly associated with milk traits in other studies. These SNPs were located on chromosome 3, 5, 6, 8, 10, 20, and X. In contrast, no association to the well-known *DGAT1* region could be detected in DSN. Further analysis is needed to make sure that these SNPs do not negatively affect other important

production traits or DSN characteristic traits such as carcass and meat, conformation, fertility, and health before using them as breeding markers. Nevertheless, the results of this study are a basis for further genetic analysis to identify genes and causal variants that affect milk traits in DSN cattle as well as in other related breeds.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repositories and accession numbers can be found below: The European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) European Nucleotide Archive (ENA) and European Variation Archive (EVA), <https://www.ebi.ac.uk/ena/browser/home> and <https://www.ebi.ac.uk/eva/>, PRJEB42513 (project) and ERZ1701738 (analyses).

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because samples were collected based on routine procedures on these farm animals. Ear tags were taken as part of the required registration procedure, blood samples were taken by a trained veterinarian to perform standard health recording. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

GB, PK, and DA designed the study. PK performed all computational and statistical analysis, interpreted the data, and

drafted the manuscript. DA helped with the statistical analysis. SK and KM provided genotypes of 61 cows. DA and GB helped draft the manuscript. All authors read and approved the final manuscript.

FUNDING

The project was supported by funds of the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the parliament of the Federal Republic of Germany *via* the Federal Office for Agriculture and Food (BLE) under the Federal Program for Ecological Farming and Other Forms of Sustainable Agriculture (Funding number: 2815NA010).

ACKNOWLEDGMENTS

We acknowledge support by the German Research Foundation (DFG) and the Open Access Publication Fund of Humboldt-Universität zu Berlin. The RBB Rinderproduktion Berlin-Brandenburg GmbH and DSN farms supported this project with their expertise in animal selection and collecting ear tags. We would like to thank Monika Reißmann for sample management and the isolation of DNA and Siham Rahmatalla for her expertise on known milk genes.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.640039/full#supplementary-material>

REFERENCES

- Abdel-Shafy, H., Bortfeldt, R. H., Reissmann, M., and Brockmann, G. A. (2018). Validating genome-wide associated signals for clinical mastitis in German Holstein cattle. *Anim. Genet.* 49, 82–85. doi: 10.1111/age.12624
- Aliloo, H., Pryce, J. E., González-Recio, O., Cocks, B. G., and Hayes, B. J. (2015). Validation of markers with non-additive effects on milk yield and fertility in Holstein and Jersey cows. *BMC Genet.* 16:89. doi: 10.1186/s12863-015-0241-9
- Alipanah, M., Kalashnikova, L., and Rodionov, G. (2007). Association of genetic variants of the prolactin gene with milk production traits in Russian Red Pied cattle. *Proc. Br. Soc. Anim. Sci.* 2007, 156–156. doi: 10.1017/s1752756200020597
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene Ontology: tool for the tool for the unification of biology. *Gene Expr.* 25, 25–29. doi: 10.1038/75556
- Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–265. doi: 10.1093/bioinformatics/bth457
- Bolormaa, S., Porto Neto, L. R., Zhang, Y. D., Bunch, R. J., Harrison, B. E., Goddard, M. E., et al. (2011). A genome-wide association study of meat and carcass traits in Australian cattle. *J. Anim. Sci.* 89, 2297–2309. doi: 10.2527/jas.2010-3138
- Bossé, R., Fumagalli, F., Jaber, M., Giros, B., Gainetdinov, R. R., Wetsel, W. C., et al. (1997). Anterior pituitary hypoplasia and dwarfism in mice lacking the dopamine transporter. *Neuron* 19, 127–138. doi: 10.1016/S0896-6273(00)80353-0
- Buitenhuis, B., Poulsen, N. A., Gebreyesus, G., and Larsen, L. B. (2016). Estimation of genetic parameters and detection of chromosomal regions affecting the major milk proteins and their post translational modifications in Danish Holstein and Danish Jersey cattle. *BMC Genet.* 17:114. doi: 10.1186/s12863-016-0421-2
- Caroli, A. M., Chessa, S., and Erhardt, G. J. (2009). Invited review: milk protein polymorphisms in cattle: effect on animal breeding and human nutrition. *J. Dairy Sci.* 92, 5335–5352. doi: 10.3168/jds.2009-2461
- Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7. doi: 10.1186/s13742-015-0047-8
- Cole, J. B., Wiggans, G. R., Ma, L., Sonstegard, T. S., Lawlor, T. J., Crooker, B. A., et al. (2011). Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. *BMC Genomics* 12:408. doi: 10.1186/1471-2164-12-408
- Dadousis, C., Biffani, S., Cipolat-Gotet, C., Nicolazzi, E. L., Rosa, G. J. M., Gianola, D., et al. (2017). Genome-wide association study for cheese yield and curd nutrient recovery in dairy cows. *J. Dairy Sci.* 100, 1259–1271. doi: 10.3168/jds.2016-11586

- Devlin, B., and Roeder, K. (1999). Genomic control for association studies. *Biometrics* 55, 997–1004. doi: 10.1111/j.0006-341X.1999.00997.x
- Doran, A. G., Berry, D. P., and Creevey, C. J. (2014). Whole genome association study identifies regions of the bovine genome and biological pathways involved in carcass trait performance in Holstein-Friesian cattle. *BMC Genomics* 15:837. doi: 10.1186/1471-2164-15-837
- Durinck, S., Spellman, P. T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/ Bioconductor package biomaRt. *Nat. Protoc.* 4, 1184–1191. doi: 10.1038/nprot.2009.97
- Dybus, A. (2002). Associations of growth hormone [GH] and prolactin [PRL] genes polymorphisms with milk production traits in Polish Black-and-White cattle. *Anim. Sci. Pap. Rep.* 20, 203–212.
- Eigler, T., and Ben-Shlomo, A. (2014). Somatostatin system: molecular mechanisms regulating anterior pituitary hormones. *J. Mol. Endocrinol.* 53, R1–R19. doi: 10.1530/JME-14-0034
- Gao, X., Starmer, J., and Martin, E. R. (2008). A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet. Epidemiol.* 32, 361–369. doi: 10.1002/gepi.20310
- Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., et al. (2002). Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Res.* 12, 222–231. doi: 10.1101/gr.224202
- Grothe, P. O. (1993). *Holstein-Friesian: Eine Rasse geht um die Welt*. Münster: Landwirtschaftsverlag.
- Hawken, R. J., Zhang, Y. D., Fortes, M. R. S., Collis, E., Barris, W. C., Corbet, N. J., et al. (2012). Genome-wide association studies of female reproduction in tropically adapted beef cattle. *J. Anim. Sci.* 90, 1398–1410. doi: 10.2527/jas.2011-4410
- Hu, Z. L., Park, C. A., and Reecy, J. M. (2019). Building a livestock genetic and genomic information knowledgebase through integrative developments of animal QTLdb and CorrDB. *Nucleic Acids Res.* 47, D701–D710. doi: 10.1093/nar/gky1084
- Jiang, J., Ma, L., Prakash, D., VanRaden, P. M., Cole, J. B., and Da, Y. (2019). A large-scale genome-wide association study in U.S. Holstein Cattle. *Front. Genet.* 10:412. doi: 10.3389/fgene.2019.00412
- Kassambara, A. (2020). *ggpubr: “ggplot2” Based Publication Ready Plots*. Available online at: <https://cran.r-project.org/package=ggpubr> (accessed September 1, 2020).
- Kolbehdari, D., Wang, Z., Grant, J. R., Murdoch, B., Prasad, A., Xiu, Z., et al. (2009). A whole genome scan to map QTL for milk production traits and somatic cell score in Canadian Holstein bulls. *J. Anim. Breed. Genet.* 126, 216–227. doi: 10.1111/j.1439-0388.2008.00793.x
- Köppe-Försthoff, J. (1967). *100 Jahre Deutsche Schwarzbuntzucht*. Hiltrup: Verband deutscher Schwarzbuntzüchter e.V.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82:26. doi: 10.18637/jss.v082.i13
- Li, Y., Gao, Y., Kim, Y.-S., Iqbal, A., and Kim, J.-J. (2017). A whole genome association study to detect additive and dominant single nucleotide polymorphisms for growth and carcass traits in Korean native cattle, Hanwoo. *Asian-Australas. J. Anim. Sci.* 30, 8–19. doi: 10.5713/ajas.16.0170
- Logarajah, S., Hunter, P., Kraman, M., Steele, D., Lakhani, S., Bobrow, L., et al. (2003). BCL-6 is expressed in breast cancer and prevents mammary epithelial differentiation. *Oncogene* 22, 5572–5578. doi: 10.1038/sj.onc.1206689
- Lu, H., and Bovenhuis, H. (2019). Genome-wide association studies for genetic effects that change during lactation in dairy cattle. *J. Dairy Sci.* 102, 7263–7276. doi: 10.3168/jds.2018-15994
- Mapholi, N. O., Maiwashe, A., Matika, O., Riggio, V., Bishop, S. C., MacNeil, M. D., et al. (2016). Genome-wide association study of tick resistance in South African Nguni cattle. *Ticks Tick Borne Dis.* 7, 487–497. doi: 10.1016/j.ttbdis.2016.02.005
- Mateescu, R. G., Garrick, D. J., and Reecy, J. M. (2017). Network analysis reveals putative genes affecting meat quality in Angus cattle. *Front. Genet.* 8:171. doi: 10.3389/fgene.2017.00171
- May, K., Schep, C., Brügemann, K., Yin, T., Strube, C., Korkuć, P., et al. (2019). Genome-wide associations and functional gene analyses for endoparasite resistance in an endangered population of native German Black Pied cattle. *BMC Genomics* 20:277. doi: 10.1186/s12864-019-5659-4
- McClure, M. C., Ramey, H. R., Rolf, M. M., McKay, S. D., Decker, J. E., Chapple, R. H., et al. (2012). Genome-wide association analysis for quantitative trait loci influencing Warner-Bratzler shear force in five taurine cattle breeds. *Anim. Genet.* 43, 662–673. doi: 10.1111/j.1365-2052.2012.02323.x
- Mckenzie, H. A., Bruce Graham, E. R., and Ponzone, R. W. (1984). Effects of milk protein genetic variants on milk yield and composition. *J. Dairy Res.* 51, 531–546. doi: 10.1017/S0022029900032854
- Meier, S., Arends, D., Korkuć, P., Neumann, G. B., and Brockmann, G. A. (2020). A genome-wide association study for clinical mastitis in the dual-purpose German Black Pied cattle breed. *J. Dairy Sci.* 103, 10289–10298. doi: 10.3168/jds.2020-18209
- Meier, S., Korkuć, P., Arends, D., and Brockmann, G. A. (2019). DNA sequence variants and protein haplotypes of casein genes in German Black Pied Cattle (DSN). *Front. Genet.* 10:1129. doi: 10.3389/fgene.2019.01129
- Meredith, B. K., Kearney, F. J., Finlay, E. K., Bradley, D. G., Fahey, A. G., Berry, D. P., et al. (2012). Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. *BMC Genet.* 13:21. doi: 10.1186/1471-2156-13-21
- Mészáros, G., Petautschnig, E., Schwarzenbacher, H., and Sölkner, J. (2015). Genomic regions influencing coat color saturation and facial markings in Fleckvieh cattle. *Anim. Genet.* 46, 65–68. doi: 10.1111/age.12249
- Nayeri, S., Sargolzaei, M., Abo-Ismael, M. K., May, N., Miller, S. P., Schenkel, F., et al. (2016). Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. *BMC Genet.* 17:75. doi: 10.1186/s12863-016-0386-1
- Nayeri, S., Schenkel, F., Fleming, A., Kroezen, V., Sargolzaei, M., Baes, C., et al. (2019). Genome-wide association analysis for β -hydroxybutyrate concentration in Milk in Holstein dairy cattle. *BMC Genet.* 20:58. doi: 10.1186/s12863-019-0761-9
- Ng-Kwai-Hang, K. F. F., Hayes, J. F. F., Moxley, J. E. E., and Monardes, H. G. G. (1984). Association of genetic variants of casein and milk serum proteins with milk, fat, and protein production by dairy cattle. *J. Dairy Sci.* 67, 835–840. doi: 10.3168/jds.S0022-0302(84)81374-0
- Ogorevc, J., Kunj, T., Razpet, A., and Dovc, P. (2009). (milk production and mastitis traits) Database of cattle candidate genes and genetic markers for milk production and mastitis. *Anim. Genet.* 40, 832–851. doi: 10.1111/j.1365-2052.2009.01921.x
- Parker Gaddis, K. L., Null, D. J., and Cole, J. B. (2016). Explorations in genome-wide association studies and network analyses with dairy cattle fertility traits. *J. Dairy Sci.* 99, 6420–6435. doi: 10.3168/jds.2015-10444
- Poulsen, N. A., Rybicka, I., Larsen, L. B., Buitenhuis, A. J., and Larsen, M. K. (2015). Short communication: genetic variation of riboflavin content in bovine milk. *J. Dairy Sci.* 98, 3496–3501. doi: 10.3168/jds.2014-8829
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- R Core Team (2018). *A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- RBB Rinderproduktion Berlin-Brandenburg GmbH. (2020). *Deutsches Schwarzbuntes Niederungsgrind – lebendes Kulturerbe*. Available online at: <https://www.rinderzucht-bb.de/zucht/dsn-genreserve/> (accessed April 1, 2020).
- Rosen, B. D., Bickhart, D. M., Schnabel, R. D., Koren, S., Elsik, C. G., Zimin, A., et al. (2018). Modernizing the bovine reference genome assembly. *Proc. World Congr. Genet. Appl. Livest. Prod.* 3:802.
- Snelling, W. M., Allan, M. F., Keele, J. W., Kuehn, L. A., McDanel, T., Smith, T. P. L., et al. (2010). Genome-wide association study of growth in crossbred beef cattle. *J. Anim. Sci.* 88, 837–848. doi: 10.2527/jas.2009-2257
- The Society for the Conservation of Old and Endangered Livestock Breeds (GEH). (2018). *Rassenbeschreibungen Rinder: Deutsches Schwarzbuntes Niederungsgrind*. Available online at: <http://www.g-e-h.de/index.php/rassebeschreibungen/34-rassekurzbeschreibungen-rinder/70-deutsches-schwarbuntes-niederungsgrind> (accessed April 1, 2020).
- Wang, Z., and Chatterjee, N. (2017). Increasing mapping precision of genomewide association studies: to genotype and impute, sequence, or both? *Genome Biol.* 18, 17–19. doi: 10.1186/s13059-017-1255-6

- Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., et al. (2016). *Create Elegant Data Visualisations Using the Grammar of Graphics*. R package ggplot2 version 3.3.2.
- Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., et al. (2020). Ensembl 2020. *Nucleic Acids Res.* 48, D682–D688. doi: 10.1093/nar/gkz966
- Zielke, L. G., Bortfeldt, R. H., Reissmann, M., Tetens, J., Thaller, G., and Brockmann, G. A. (2013). Impact of variation at the FTO locus on milk fat yield in holstein dairy cattle. *PLoS One* 8:e63406. doi: 10.1371/journal.pone.0063406
- Zielke, L. G., Bortfeldt, R. H., Tetens, J., and Brockmann, G. A. (2011). BDNF contributes to the genetic variance of milk fat yield in German Holstein cattle. *Front. Genet.* 2:16. doi: 10.3389/fgene.2011.00016

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Korkuć, Arends, May, König and Brockmann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.